Chapter 4

Camera Calibration

Recovering 3D structure from images becomes a simpler problem when the images are taken with *calibrated* cameras. For our purposes, a camera is said to be *calibrated* if the mapping between image coordinates and directions relative to the camera center are known. However, the position of the camera in space (i.e. its translation and rotation with respect to world coordinates) is not necessarily known.

4.1 The perspective model

For an ideal pinhole camera delivering a true perspective image, this mapping can be characterized completely by just five numbers, called the *intrinsic parameters* of the camera. In contrast, a camera's *extrinsic parameters* represent its location and rotation in space. The five intrinsic camera parameters are:

1. The x-coordinate of the the center of projection, in pixels (u_0)

2. The y-coordinate of the the center of projection, in pixels (v_0)

3. The focal length, in pixels (f)

4. The aspect ratio (a)

5. The angle between the optical axes (c)

An excellent presentation of the algebraic and matrix representations of perspective cameras may be found in [13].

The calibration of real cameras is often approximated with such a five-parameter mapping, though rarely is such an approximation accurate to the pixel. The next section will discuss how real images differ from true perspective and a procedure for correcting for the difference. While several methods have been presented to determine the the intrinsic parameters of a camera from images it has taken, approximate values for these parameters can often be assumed *a priori*.

For most images taken with standard lenses, the **center of projection** is at or near the coordinate center of the image. However, small but significant shifts are often introduced in the image recording or digitizing process. Such an image shift has the most impact on camera calibration for lenses with shorter focal lengths. Images that have been cropped may also have centers of projection far from the image center.

Old-fashioned bellows cameras, and modern cameras with tilt-shift lenses, can be used to place the center of projection far from the center of the image. This is most often used in architectural photography, when the photographer places the film plane vertically and shifts the lens upward until the top of the building projects onto the film area. The result is that vertical lines remain parallel rather than converging toward the top of the picture, which would happen if the photographer had simply rotated the camera upwards. To extract useful geometric information out of such images, it is necessary to compute the center of projection.

The **focal length** of the camera in pixels can be estimated by dividing the marked focal length of the camera lens by the width of the image on the imaging surface (the film or CCD array), and then multiplying by the width of the final image in pixels. For example, the images taken with a 35mm film camera are 36mm wide and 24mm high[1]. Thus an image taken with a 50mm lens and digitized at 768 by 512 pixels would have an approximate focal length of $(50/36) \times 768 = 1067$ pixels. The reason this is approximate is that the digitization process typically crops out some of the original image, which increases the observed focal length slightly. Of course, zoom lenses are variable in focal length so this procedure may only be applied if the lens was known to be fully extended or retracted.

It should be noted that most prime[2] lenses actually change in focal length depending on the distance at which they are focussed. This means that images taken with the same lens on the same camera may exhibit different focal lengths, and thus need separate camera calibrations. The easiest solution to this problem is to fix the focus of the lens at infinity and use a small enough aperture small to image the closest objects in the scene in focus. Another solution is to use *telecentric* lenses, whose focal length is independent of focus. A procedure for converting certain regular lenses to telecentric ones is presented in [55].

The **aspect ratio** for images taken with real cameras with radially symmetric lens elements is 1.0, although recording and digitizing processes can change this. The Kodak PhotoCD process for digitizing film maintains a unit aspect ratio. Some motion-picture cameras used to film wide screen features use non-radially symmetric optics to squeeze a wide image into a relatively narrow frame; these images are then expanded during projection. In this case, the aspect ratio is closer to 2.0.

---

[1] 35mm refers to the height of the entire film strip, including the sprocket holes

[2] A prime lens is a lens with a fixed focal length, as opposed to a zoom lens, which is variable in focal length.

Finally, for all practical cases of images acquired with real cameras and digitized with standard equipment, the **angle between the optical axes** is 90 degrees.

4.2 How real cameras deviate from the pinhole model

Real cameras deviate from the pinhole model in several respects. First, in order to collect enough light to expose the film, light is gathered across the entire surface of the lens. The most noticeable effect of this is that only a particular surface in space, called the *focal plane*[3], will be in perfect focus. In terms of camera calibration, each image point corresponds not to a single ray from the camera center, but to a set of rays from across the front of the lens all converging on a particular point on the focal plane. Fortunately, the effects of this area sampling can be made negligible by using a suitably small camera aperture.

The second, and most significant effect, is lens distortion. Because of various constraints in the lens manufacturing process, straight lines in the world imaged through real lenses generally become somewhat curved on the image plane. However, since each lens element is radially symmetric, and the elements are typically placed with high precision on the same optical axis, this distortion is almost always radially symmetric, and is referred to as *radial lens distortion*. Radial distortion that causes the image to bulge toward the center is called *barrel* distortion, and distortion that causes the image to shrink toward the center is called *pincushion* distortion. Some lenses actually exhibit both properties at different scales.

To correct for radial distortion, one needs to recover the center of the distortion (c_x, c_y), usually consistent with the center of projection of the image, and a radial transformation function

---

[3]Although called the focal plane, this surface is generally slightly curved for real lenses

that remaps radii from the center such that straight lines stay straight:

$$r' = F(r) \tag{4.1}$$

Usually, the radial distortion function is modeled as r multiplied by an even polynomial of

the form:

$$F(r) = r(1 + k_1 r^2 + k_2 r^4 + ...) \tag{4.2}$$

The multiplying polynomial is even in order to ensure that the distortion is C^∞ continu-

ous at the center of distortion, and the first coefficient is chosen to be unity so that the original and

undistorted images agree in scale at the center of distortion. These coefficients can be determined

by measuring the curvature of putatively straight lines in images. Such a method will be presented

in the next section.

The distortion patterns of cameras with imperfectly ground or imperfectly aligned optics

may not be radially symmetric, in which case it is necessary to perform a more general distortion

correction.

Another deviation from the pinhole model is that in film cameras the film plane can deviate

significantly from being a true plane. The plate at the back of the camera may not be perfectly flat,

or the film may not lie firmly against it. Also, many film digitization methods do not ensure that the

film is perfectly flat during the scanning process. These effects, which we collectively refer to as *film*

flop, cause subtle deformations in the image. Since some of the deformations are different for each

photograph, they cannot be corrected for beforehand through camera calibration. Digitial cameras,

which have precisely flat and rectilinear imaging arrays, are generally not susceptible to this sort of

distortion.

A final, and particularly insidious deviation from the pinhole camera model is that the imaged rays do not necessarily intersect at a point. As a result, there need not be a mathematically precise principal point, or nodal point for a real lens, as illustrated in Fig. 4.1. As a result, it is impossible to say with complete accuracy that a particular image was taken from a particular location in space; each pixel must be treated as its own separate ray. Although this effect is most noticeable in extreme wide-angle lenses, the locus of convergence is almost always small enough to be treated as a point, especially when the objects being imaged are large with respect to the locus of convergence.
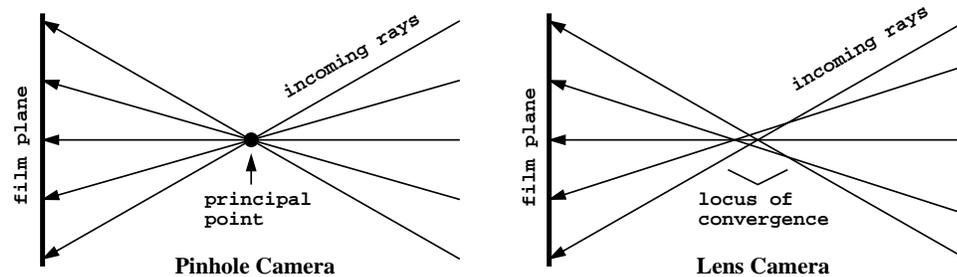
Figure 4.1: In a pinhole camera, all the imaged rays must pass though the pinhole, which effectively becomes the mathematical location of the camera. In a real camera with a real lens, the imaged rays need not all intersect at a point. Although this effect is usually insignificant, to treat it correctly would complicate the problems of camera calibration and 3D reconstruction considerably.

Considerable work has been done in both photogrammetry and computer vision to calibrate cameras and lenses for both their perspective intrinsic parameters and their distortion patterns. Some successful methods include [52], [12], and [11]. While there has been recent progress in the use of uncalibrated views for 3D reconstruction [14], this method does not consider non-perspective camera distortion which prevents high-precision results for images taken through real lenses. In our work, we have found camera calibration to be a straightforward process that considerably simplifies the problem of 3D reconstruction. The next section presents the camera calibration process used for

our project.

4.3   Our calibration method

Our calibration method uses two calibration objects. For each camera/lens configuration used in the reconstruction project, a few photographs of each calibration object are taken. The first calibration object (Fig. 4.2) is a flat checkerboard pattern, and is used to recover the pattern of radial distortion from the images. The second object (Fig. 4.8) is two planes with rectangular patterns set at a 90 degree angle to each other, and is used to recover the intrinsic perspective parameters of the camera.

4.4   Determining the radial distortion coefficients

The first part of the calibration process is to determine an image coordinate remapping that causes images taken by the camera to be true perspective images, that is, straight lines in the world project as straight lines in the image. The procedure makes use of one or several images with many known straight lines in it. Architectural scenes are usually a rich source of straight lines, but for most of the work in this thesis we used pictures of the checkerboard pattern shown below (Fig. 4.2) to determine the radial lens distortion. The checkerboard pattern is a natural choice since straight lines with easily localized endpoints and interior points can be found in several orientations (horizontal, vertical, and various diagonals) throughout the image plane.

The checkerboard pattern also has the desirable property that its corners are localizable independent of the linearity of the image response. That is, applying a nonlinear monotonic function to the intensity values of the checkerboard image, such as gamma correction, does not affect corner

localization. As a counterexample, this is not the case for the corners of a white square on a black background. If the image is blurred somewhat, changing the image gamma will cause the square to shrink or enlarge, which will affect corner localization.
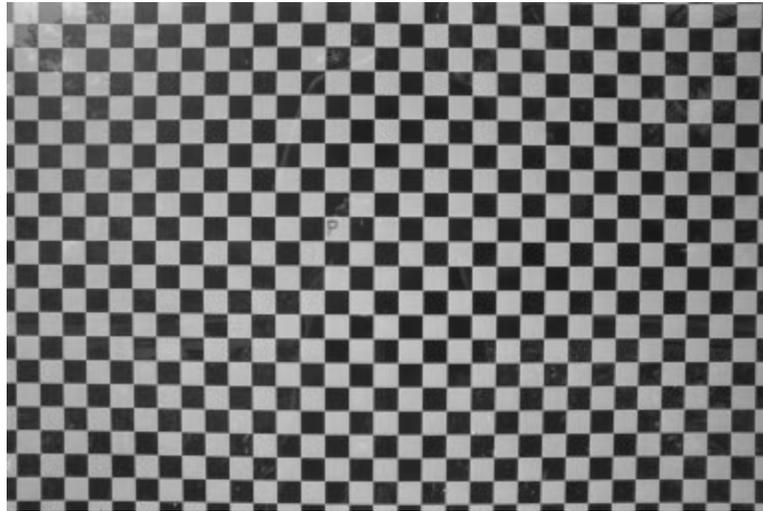
Figure 4.2: Original image of a calibration checkerboard pattern, taken with a Canon 24mm EF lens. The straight lines in several orientations throughout this image are used to determine the pattern of radial lens distortion. The letter "P" in the center is used to record the orientation of the grid with respect to the camera.

The pattern in Fig. 4.2 was photographed with a 24mm lens on a Canon EOS Elan camera. Since this lens, like most, changes its internal configuration depending on the distance it is focussed at, it is possible that its pattern of radial distortion could be different depending on where it is focussed. Thus, care was taken to focus the lens at infinity and to reduce the aperture until the image was adequately sharp. Clearly, this procedure works only when the calibration object is far enough from the camera to be brought into focus via a small aperture. Since wide-angle lenses generally have large depths of field, this was not a problem for the 24mm lens with a 50cm high calibration grid. However, the depth of field of a 200mm lens was too shallow to focus the object even when fully stopped down — a larger calibration object, placed further away, was called for.

The pattern of radial distortion in Fig. 4.2 may be too subtle to be seen directly, so I have developed a procedure for more easily visualizing lens distortion in checkerboard test images. First, a simple Sobel edge detector is run on the image to produce the image shown in 4.3.
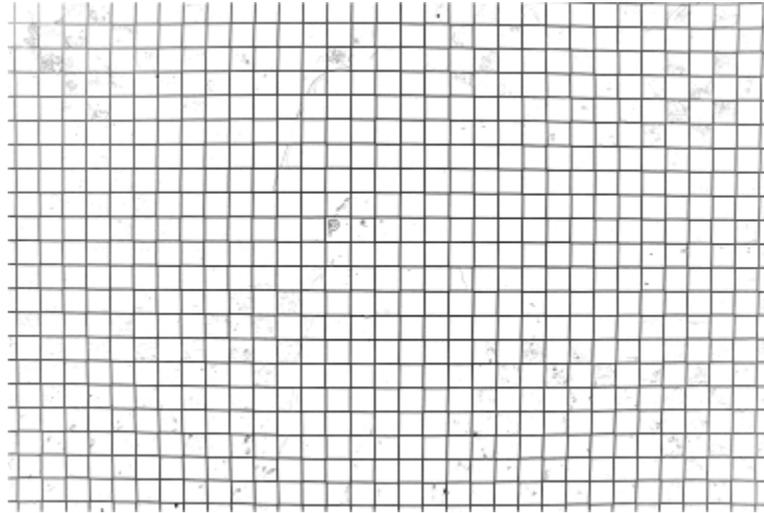
Figure 4.3: The edges of the checkerboard pattern found by using a simple Sobel edge detector. (Shown in reverse video)

The pattern of distortion can now be made evident to a human observer by shrinking this edge image in either the horizontal or the vertical direction by an extreme amount. Fig. 4.4 shows this edge image shrunk in both the vertical and horizontal directions by a factor of 50. In the case of this 24mm lens, we can see that lines passing through the center of the image stay straight, as do the vertical lines at the extreme left and right of the image. Lines which lie at intermediate distances from the center of the image are bowed. The bottom image, resulting from shrinking the image in the horizontal direction and rotating by 90 degrees, shows that this bowing is actually not convex. We will see this represented in the radial distortion coefficients as a positive k_1 and a negative k_2.

The choice of the checkerboard pattern makes it possible to automatically localizing image points. Image points can be easily localized by first convolving the image with the filter in Table

Figure 4.4: The results of changing the aspect ratio of the image in Fig. 4.3 by a factor of 50 in both the horizontal (top) and vertical (bottom, rotated 90 degrees) directions. This extreme change in aspect ratio makes it possible for a human observer to readily examine the pattern of lens distortion.

-1	-1	-1	0	1	1	1
-1	-1	-1	0	1	1	1
-1	-1	-1	0	1	1	1
0	0	0	0	0	0	0
1	1	1	0	-1	-1	-1
1	1	1	0	-1	-1	-1
1	1	1	0	-1	-1	-1

Table 4.1: A 7×7 convolution filter that detects corners of the checkerboard pattern.

4.4. Since this filter itself resembles a checkerboard pattern, it gives a strong response (positive or negative, depending on which type of corner) when centered over a checkerboard corner. Taking the absolute value of the filter output produces an image where the checkerboard corners appear as white dots, as in Fig. 4.5.

Localizing a particular checkerboard corner after the filter convolution is easily accomplished by locating the point of maximum filter response. Sub-pixel accuracy can be obtained by examining the filter responses at pixels neighboring the pixel of maximum response, fitting these responses with an upside-down paraboloid, and calculating the location of the global maximum of the paraboloid.

The localized checkerboard corners provide many sets of points which are collinear in the world. In fact, these sets of points can be found in many orientations, including horizontal, vertical, and diagonal. However, because of lens distortion, these points will in general not be precisely collinear in the image. For any such set of points, one can quantify its deviation from linear by fitting
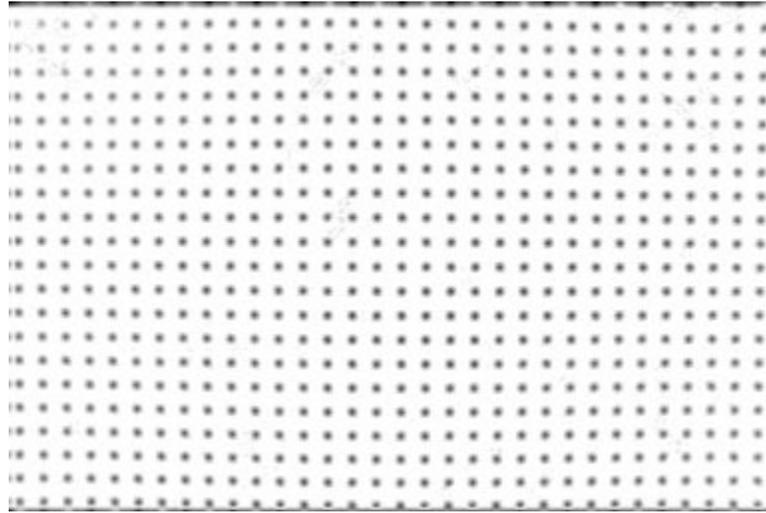
Figure 4.5: The results of convolving the checkerboard image with the filter in Table 4.4, and taking the absolute value of the filter outputs. Both corners that are white at the upper-left and black at the upper left become easily detectable dots. (Shown in reverse video)

a line to the set of points in a least squares sense and summing the squared distances of the points

from the line. In the distortion correction method described here, triples of world-collinear points are

used to measure the lens distortion. The amount of error contributed by a triple of points is shown

in Fig. 4.6.
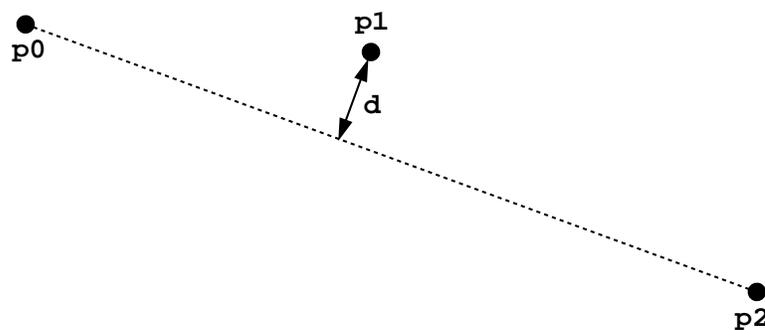
Figure 4.6: The distortion error function for a single line of three world-collinear points. The error is the distance d between the middle point p_1 from the line connecting the endpoints $p0$ and $p2$. This error is summed over many triples of world-collinear points to form an objective function, which is then optimized to determine the radial distortion coefficients of the lens.

The errors for many triples of world-collinear points throughout the image are summed to

produce an objective function $\mathbf{O_d}$ that measures the extent to which the lens deviates from the true pinhole model. Applying a radial distortion pattern with parameters $(c_x, c_y, k_1, k_2, k_3)$ to the coordinates of the localized point triples will change the value of $\mathbf{O_d}$, and when the best values for these parameters are chosen, the objective function will be at its global minimum. Thus, the radial distortion parameters can be computed by finding the minimum of $\mathbf{O_d}$. For this work the minimum was found for a variety of lenses using the `fminu` function of the MATLAB numerical analysis package. For the particular 1536×1024 image in 4.2, the parameters computed were:

$$c_x = 770.5 \tag{4.3}$$

$$c_y = 506.0 \tag{4.4}$$

$$k_1 = 9.46804 \times 10^{-8} \tag{4.5}$$

$$k_2 = -7.19742 \times 10^{-14} \tag{4.6}$$

Note that the center of distortion, $(770.5, 506.0)$ is near but not at the center of the image $(768, 512)$. The fact that the two distortion coefficients k_1 and k_2 are opposite in sign models the wavy, non-convex nature of some of the distorted lines seen in Fig. 4.4.

Once the distortion parameters are solved for, it is possible to undistort any image taken with the same lens as the calibration images so that straight lines in the world image to straight lines on the image plane. In this work, the undistortion process could be performed without loss of image quality because the PhotoCD images were available in higher resolutions than those that were used in the reconstruction software. Specifically, images at 1536×1024 pixels were undistorted using sub-pixel bilinear interpolation and then filtered down to 768×512 pixels for use in the software, making any loss of image quality due to resampling negligible. Note that performing this resampling requires

the construction of a backward coordinate lookup function from undistorted to distorted image co-ordinates, which requires finding the inverse of the distortion polynomial in Equation 4.2. Since this is difficult to perform analytically, in this work the inverse function was inverted numercaly.

As a test of the radial distortion calibration, one can undistort the calibration images themselves and see if the original straight lines become straight. Fig. 4.7 shows the results of undistorting the original checkerboard image, and just below with edge detection and shrinking as in Figs. 4.3 and 4.4 to better reveal the straightness of the lines in the image.
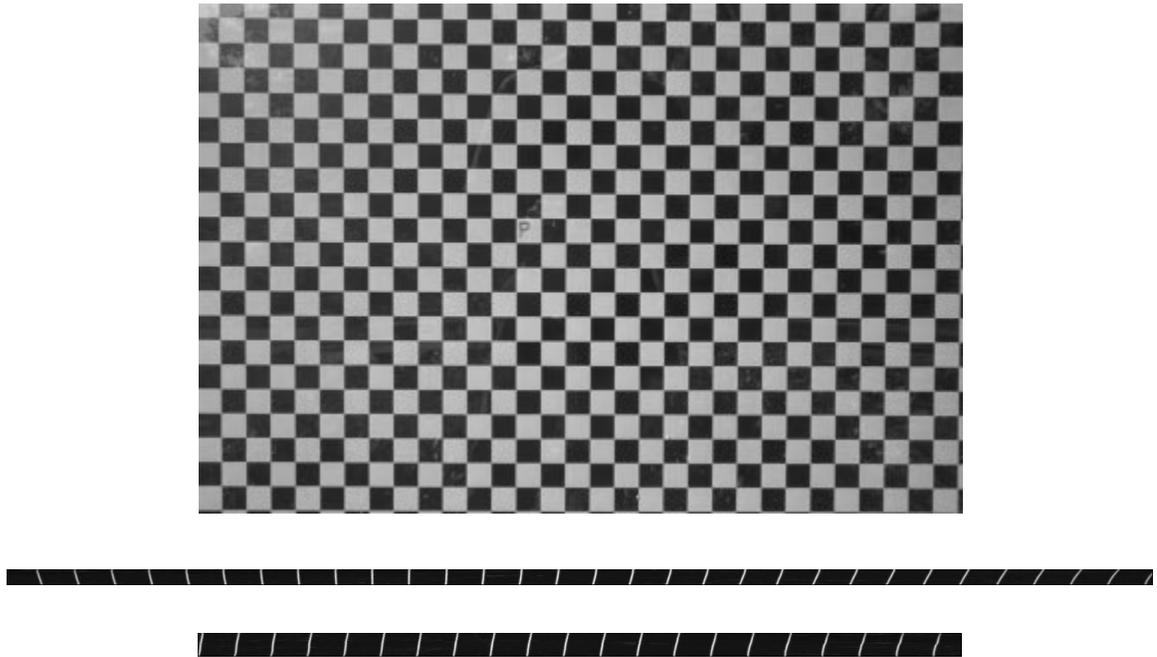
Figure 4.7: **Top:** The results of solving for the radial distortion parameters of the 24mm lens based on the points in Fig. 4.5, and unwarping the original grid image (Fig. 4.2) and running edge detection on it. **Bottom:** The two bottom images, scaled by a factor of fifty in either direction, help verify that with the distortion correction, the lines are now straight. Compare to the curved lines in 4.4. Note that the lines are not parallel; this is because the camera's film plane was not placed exactly parallel to the plane of the grid. It is a strength of this method that such alignment is not necessary.

4.5 Determining the intrinsic parameters

Once the distortion pattern of a lens is known, we can use any image taken with that lens, undistort it, and then have an image in which straight lines in the world will project to straight lines in the image. As a result, the projection is now a true perspective projection, and it becomes possible to characterize the lens in terms of its five intrinsic parameters (see Sec. 4.1).

For this project, we used a calibration process provided to us by Q.T. Luong [11]. In this method, an image of a calibration object, shown in Fig. 4.8, is used to determine the intrinsic parameters. The computer knows the geometry of the model *a priori*, and since the model has sufficient 3D structure, the computer can solve for the eleven-degree-of-freedom projection matrix that would give rise to the image of the object. This matrix is then factored into the camera's six extrinsic parameters (translation and rotation of the camera relative to the object) and the five intrinsic parameters. In practice, solving for the 4×3 projection matrix is done with a nonlinear optimization over its twelve elements, and the process is given an initial estimate by the user.
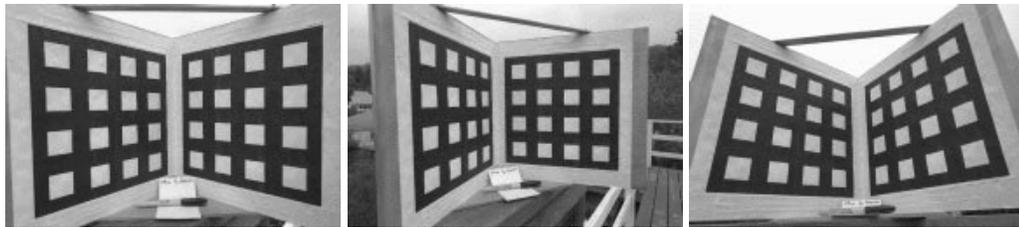
Figure 4.8: Q.T. Luong's calibration object, photographed at several orientations. The three photographs of the object were used to recover the intrinsic parameters of the 24mm Canon EOS lens used to take the photographs. Before solving for the perspective projection parameters, the lens' radial distortion pattern was modeled separately using a checkerboard grid object (Fig. 4.2). As a result, the lens calibration was far more accurate.

A more reliable estimate of the intrinsic camera parameters can be obtained by photographing the grid in several orientations with respect to each camera lens. Fig. 4.8 shows the object at

Before undistorting	α_u	α_v	f	a
Image 1	559.603	557.444	389.686	1.019
Image 2	548.419	547.001	400.529	1.056
Image 3	555.370	551.893	383.531	1.073
After undistorting	α_u	α_v	f	a
Image 1	531.607	530.964	387.109	1.017
Image 2	529.178	528.188	387.173	1.018
Image 3	530.988	529.954	387.930	1.029

Table 4.2: Computed intrinsic parameters for the three images of the calibration object in Fig. 4.8, with and without first solving and correcting for radial distortion. The fifth intrinsic parameter, the angle between the optical axes c, is not shown since it was negligibly different from ninety degrees. Note that the parameters are far more consistent with each other after correcting for radial distortion. Without the correction, the distortion introduces different errors into the calibration depending on the position and orientation of the calibration object.

three different orientations with a particular 24mm lens. The intrinsic parameters were solved for separately using each orientation, and these results were averaged to obtain the final estimate of the parameters. The importance of first solving for the pattern of radial distortion was well illustrated by an attempt to solve for the intrinsic parameters in the three separate images with and without distortion correction. Table 4.5 shows that the parameters derived from the three images were suitably consistent with each other using distortion correction, and much less consistent without distortion correction.

4.6 Working with uncalibrated images

While camera calibration is a simple and straightforward process that can simplify photogrammetry considerably, some very attractive applications require the use of images taken with uncalibrated cameras. (One such application is described in Chapter 8). Specifically, photographs exist for many architectural scenes that have since been modified or destroyed. As an example, Berkeley's original campus designed by John Galen Howard in the late 1800's featured half a dozen Victorian

brick structures, of which only South Hall remains. Nonetheless, hundreds of photos of North Hall,

Bacon Hall, and the rest of the original campus sit in the university's archives (for an example, see

Fig. 4.9). To use these photographs, it is necessary to determine the camera parameters without pic-

tures of calibration objects.

Figure 4.9: The original Berkeley campus in the late 1800's, looking to the East. Of the four build-
ings, only South Hall on the right remains. The ability to reconstruct buildings long since destroyed
is a principal attraction of modeling architecture from photographs.

Fortunately, for architectural scenes, the buildings themselves are often serviceable cali-

bration objects. Straight lines, prevalent in architecture, can be used to determine radial lens dis-

tortion directly from the original photographs using the same method presented in Sec. 4.4. Shawn

Becker [4] presents another method of solving for radial distortion by observing sets of parallel lines,

also prevalent in architectural images.

The perspective intrinsic camera parameters can often be determined directly by observing

the vanishing points of orthogonal sets of parallel lines. If the aspect ratio a can be assumed to be

one, and the angle between the optical axes can be assumed to 90 degrees, then the remaining cam-

era parameters (center of projection and focal length) can be determined by observing the vanishing

points of three mutually orthogonal sets of lines. Geometrically, one simply needs to construct the

triangle connecting the three vanishing points on the image plane, and then intersect the image with

the corner of a cube such that each side of this triangle is coincident with a different face of the cube.

The corner of the cube will then be at the original camera center; the center of projection is obtained by dropping a perpendicular from the camera center to the image plane, and the focal length is the length of this perpendicular. Another discussion of related calibration techniques may be found in [51].

Often, photos of since-demolished or modified buildings show structures that still exist, as in Fig. 4.9. In these cases, it is possible to acquire calibrated images to reconstruct the existing architecture and then use these dimensions to recover the historic camera parameters. This technique, in conjunction with the vanishing-point technique just described, was used to determine the intrinsic camera parameters of the historic photographs used in the Rouen Revisited art installation (Sec. 8).

Some researchers [14] have explored the mathematical theory and performed experiments to recover structure from uncalibrated views. As mentioned, these techniques are not able to directly solve for radial distortion effects, and only recover structure up to an arbitrary projective transformation. Nonetheless, the techniques show that recovering some intrinsic camera information from uncalibrated views can be done implicitly as part of the structure recovery process.

Section 5 will describe our optimization technique for recovering building structure and extrinsic camera parameters from calibrated photographs. A final method of making use of uncalibrated views would be to include the intrinsic camera parameters, as well as the radial distortion coefficients, in the optimization. Further work would need to be done to use uncalibrated views in our current initial estimate generation method. Also, one should be wary of the possibilities of obtaining ill-conditioned or ambiguous solutions in situations where unconstrained descriptions of camera geometry need to be obtained at the same time as the structure of the scene.